(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 August 2003 (28.08.2003)

PCT

(10) International Publication Number
WO 03/070880 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US03/01800

(22) International Filing Date: 22 January 2003 (22.01.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/057,582 23 January 2002 (23.01.2002) US

(71) Applicant *(for all designated States except US)*: **WISCONSIN ALUMNI RESEARCH FOUNDATION** [US/US]; 614 Walnut Street, P.O. Box 7365, Madison, WI 53707-7365 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BLATTNER, Frederick, R.** [/]; 1547 Jefferson Street, Madison, WI 53711 (US). **POSFAI, Gyorgy** [/]; Majus 1 u. 74, H-6727 Szeged (HU). **HERRING, Christopher, D.** [/]; 32 Oakbridge Court, Madison, WI 53717 (US). **PLUNKETT, Guy** [/]; 1613 Gilbert Road, Madison, WI 53711 (US). **GLASNER, Jeremy, D.** [/]; 1102 East Johnson Street, Madison, WI 53706 (US). **TWOSE, Trevor, Martin** [/]; 1506 Chandler Street, Madison, WI 53711 (US).

(74) Agent: **SEAY, Nicholas, J.**; Quarles & Brady LLP, P.O. Box 2113, Madison, WI 53701-21113 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/070880 A2

(54) Title: BACTERIA WITH REDUCED GENOME

(57) Abstract: The present invention provides a bacterium having a genome that is genetically engineered to be at least 2 to about 20% smaller than the genome of its native parent strain. A bacterium with a smaller genome can produce a commercial product more efficiently. The present invention also provides methods for deleting genes and other DNA sequences from a bacterial genome. The methods provide precise deletions and seldom introduces mutations to the genomic DNA sequences around the deletion sites. Thus, the methods can be used to generate a series of deletions in a bacterium without increasing the possibility of undesired homologous recombination within the genome. In addition, some of the methods provided by the present invention can also be used for replacing a region of a bacterial genome with a desired DNA sequence.

BACTERIA WITH REDUCED GENOME

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation-in-part of U.S. Patent Application serial no. 10/057,582, filed 23 January 2002 and U.S. Provisional Application serial no. 60/409,080, filed 6 September 2002, both of which are incorporated herein by reference in their entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] This invention was made with United States government support awarded by the following agency:

NIH GM35682

The United States has certain rights in this invention.

BACKGROUND OF THE INVENTION

[0003] Bacteria have been used to produce a wide range of commercial products. For example, many *Streptomyces* strains and *Bacillus* strains have been used to produce antibiotics; *Pseudomonas denitrificans* and many *Propionibacterium* strains have been used to produce vitamin B12; some other bacteria have been used to produce vitamin Riboflavin; *Brevibacterium flavum* and *Corynebacterium glutamicum* have been used to produce lysine and glutamic acid, respectively, as food additives; other bacteria have been used to produce other amino acids used as food additives; *Alcaligenes eutrophas* has been used to produce biodegradable microbial plastics; and many *Acetobacter* and *Gluconobacter* strains have been used to produce vinegar. More recently, it has become common for bacteria, such as *Escherichia coli* (*E. coli*), to be genetically engineered and used as host cells for the production of biological reagents, such as proteins and nucleic acids, in laboratory as well as industrial settings. The pharmaceutical industry supports several examples of successful products which are human proteins which are manufactured in *E. coli* cultures cultivated in a fermenter.

[0004] It is not an uncommon occurrence for normal bacterial proteins to adversely affect the production or the purification of a desired protein product from an engineered bacteria. For example, when *E. coli* bacteria are used as host cells to generate a large quantity of a desired

product encoded by a gene that is introduced into the host cells by a plasmid, certain normal *E. coli* gene products can interfere with the introduction and maintenance of plasmid DNA. More significantly, because of the economies of bacterial culture in making proteins in bacteria, often the cost of purification of a recombinant protein can be more than the cost of production, and some of the natural proteins produced by the bacterial host are sensitive purification problems. Further, many bacterial strains produce toxins that must be purified away from the target protein being produced and some strains can produce, by coincidence, native proteins that are close in size to the target protein, thereby making size separation not available for the purification process.

[0005] Also, however, the genome of a bacteria used in a fermenter to produce a recombinant protein includes many unnecessary genes. A bacteria living in a natural environment has many condition responsive genes to provide mechanisms for surviving difficult environmental conditions of temperature, stress or lack of food source. Bacteria living in a fermentation tank do not have these problems and hence do not require these condition responsive genes. The bacterial host spends metabolic energy each multiplication cycle replicating these genes. Thus the unnecessary genes and the unneeded proteins, produced by a bacterial host used for production of recombinant protein, result is a lack of efficiencies in the system that could be improved upon.

[0006] It is not terribly difficult to make deletions in the genome of a microorganism. One can perform random deletion studies in organisms by simply deleting genomic regions to study what traits of the organism are lost by the deleted genes. It is more difficult, however, to make targeted deletions of specific regions of genomic DNA and more difficult still if one of the objectives of the method is to leave no inserted DNA, here termed a "scar," behind in the organism after the deletion. If regions of inserted DNA, i.e. scars, are left behind after a genomic deletion procedure, those regions can be the locations for unwanted recombination events that could excise from the genome regions that are desirable or engender genome rearrangements. In building a series of multiple deletions, scars left behind in previous steps could become artifactual targets for succeeding steps of deletion. This is especially so when the method is used repeatedly to generate a series of deletions from the genome. In other words, the organism becomes by the deletion process genetically unstable if inserted DNA is left behind.

BRIEF SUMMARY OF THE INVENTION

[0007] The present invention provides methods for reducing the genome of an organism preferably without leaving scars in the genome.

[0008] In one embodiment, the present invention provides a bacterium having a genome that is genetically engineered to be at least two percent (2%) to twenty percent (20%) smaller than the genome of its native parent strain. Preferably, the genome is at least seven percent (7%) smaller than the genome of the native parent. More preferably, the genome is eight percent (8%) to fourteen percent (14%) to twenty percent (20%) smaller than the genome of its native parent strain. When used to produce a product, a bacterium with a smaller genome can have one or more of the following advantages. One, the production process can be more efficient either in terms of resource consumption or in terms of production speed, ultimate yield percent or all three. Two, the product purification process can be simplified or purer products can be made. Three, a product that cannot be produced before due to native protein interference can be produced. Four, the yield per cell of the desired product may be increased.

[0009] The present invention is also directed to an organism, preferably a bacterium, engineered to have a "clean genome", *i.e.*, lacking, for example, genetic material such as certain genes unnecessary for growth and metabolism of the bacteria, insertion sequences (transposable element), pseudogenes, prophage, endogenous restriction-modification genes, pathogenicity genes, toxin genes, fimbrial genes, periplasmic protein genes, invasin genes, sequences of unknown function and sequences not found in common between two strains of the same native parental species of bacterium. Other DNA sequences that are not required for cell survival and production of certain proteins in culture can be deleted. The reduced genome bacteria of the present invention may be viewed as a basic genetic framework to which may be added a myriad of genetic elements for expression of useful products as well as genetic control elements which offers an unprecedented opportunity to fine tune or optimize the expression of the desired product.

[0010] The present invention also provides materials and methods for targeted deletion of genes and other DNA sequences from a bacterial genome without leaving any residual DNA from the manipulation (scarless deletion). Since the methods of the present invention seldom introduce mutations or leave residual DNA in the genomic DNA sequences around deletion sites, the methods can be used to generate a series of deletions in a bacterium without increasing the possibility of undesired homologous recombination within the genome. Some of these methods are also useful for making similar deletions, for example, in bacteriophage, native plasmids and the like, as well as in higher organisms, such as mammals and plants.

[0011] The first deletion method is linear DNA-based. To perform the process, first, a linear DNA construct is provided in a bacterium and a region of the bacterial genome is replaced

by the linear DNA construct through homologous recombination aided by a system residing in the bacterium that can increase the frequency of homologous recombination. Next, a separate gene previously introduced into the bacterium expresses a sequence-specific nuclease to cut the bacterial genome at a unique recognition site located on the linear DNA construct. Then, a DNA sequence engineered to contain DNA homologous to a target in the genomic DNA at one end of the linear DNA construct undergoes homologous recombination with a similar genomic DNA sequence located close to the other end of the linear DNA construct. The net result is a precise deletion of a region of the genome.

[0012] The second method is also linear DNA-based. Two DNA sequences, one of which is identical to a sequence that flanks one end of a bacterial genome region to be deleted and the other of which is identical to a sequence that flanks the other end of the bacterial genome region to be deleted, are engineered into a vector in which the two sequences are located next to each other. At least one sequence-specific nuclease recognition site is also engineered into the vector on one side of the two sequences. The vector is introduced into a bacterium and a linear DNA is generated inside the bacterium by expressing inside the bacterium a nuclease that recognizes the sequence-specific nuclease recognition site and cuts the vector therein. The linear DNA undergoes homologous recombination with the bacterial genome aided by a system residing in the bacterium to increase the frequency of homologous recombination. A bacterium with a targeted deletion free of residual artifactual in its genome is thus produced.

[0013] The second method described above can also be used to replace a selected region of a bacterial genome with a desired DNA sequence. In this case, a desired DNA sequence that can undergo homologous recombination with and hence replace the selected region is engineered into the vector. All other aspects are the same as for deleting a targeted region.

[0014] The third method is suicide plasmid-based. The specific plasmid used in this method contains an origin of replication controlled by a promoter and a selectable marker, such as an antibiotic resistance gene. To delete a targeted region of a bacterial genome, a DNA insert that contains two DNA sequences located right next to each other, one of which is identical to a sequence that flanks one end of a bacterial genome region to be deleted and the other of which is identical to a sequence that flanks the other end of the bacterial genome region, is inserted into the plasmid. The plasmid is then introduced into the bacteria and integrated into the bacterial genome. Next, the promoter is activated to induce replication from the ectopic origin introduced into the bacterial genome so that recombination events are selected. In many bacteria, the

recombination events will result in a precise deletion of the targeted region of the bacterial genome and these bacteria can be identified. An alternative way to select for recombination events is to engineer a recognition site of a sequence-specific nuclease into the specific plasmid and cut the bacterial genome with the sequence-specific nuclease after the plasmid has integrated into the bacterial genome.

[0015] The suicide plasmid-based method described above can also be used to replace a selected region of a bacterial genome with a desired DNA sequence. In this case, a DNA insert that contains a desired DNA sequence that can undergo homologous recombination with and hence replace the selected region is inserted into the plasmid. All other aspects are the same as for deleting a targeted region.

[0016] The methods of the present invention are useful *inter alia* for engineering reduced genome bacteria for the production of recombinant gene products. Such engineered bacteria allow improved production of such proteins by increasing the efficiency of production and yield of the desired gene product as well as allowing more efficient purification of the product by virtue of the elimination of unnecessary bacterial gene products. A preferred reduced genome bacteria of the present invention is a bacteria from which one or more native genes encoding periplasmic proteins and/or membrane proteins have been deleted.

[0017] The present invention is also directed to DNAs and vectors used for carrying out the methods of the present invention, methods for preparing the DNAs and to kits containing vials which vials contain one or more DNAs or vectors of the present invention and optionally suitable buffers, primers, endonucleases, nucleotides, and polymerases.

[0018] The present invention is also directed to live vaccines comprising a reduced genome bacterium of the present invention or comprising a reduced genome bacterium of the present invention into which is introduced DNA encoding antigenic determinants of pathogenic organisms operably associated with expression control sequences which allow the expression of said antigenic determinants. Also within the scope of the present invention is a live vaccine comprising a reduced genome bacterium of the present invention in to which has been introduced a DNA, derived from a pathogenic organism and optionally having an origin of replication, said live vaccine being capable of inducing an enhanced immune response in a hose against a pathogenic organism. The said DNA is preferably methylated at a methylation site. The invention is also directed to a live vaccine produced from a pathogenic organism by deleting from the

genome of that organism the genes responsible for pathogenicity while retaining other antigenic determinants.

[0019] Other objects, features and advantages of the invention will become apparent upon consideration of the following detailed description.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0020] Fig. 1 shows positions of the genes and other DNA sequences on *E. coli* K-12 bacterial genome that were candidates for deletion as black and lighter hatched boxes on the outermost ring.

[0021] Fig. 2 illustrates a specific example of a linear DNA-based scarless genetic modification method of the present invention.

[0022] Fig. 3 illustrates a specific example of another linear DNA-based method of the present invention.

[0023] Fig. 4 shows a mutagenesis plasmid that can be used in the linear DNA-based method illustrated in Fig. 3.

[0024] Fig. 5A-C illustrates a specific example of a suicide plasmid-based method of the present invention.

[0025] Fig. 6 shows three plasmids that can be used in the suicide plasmid-based method illustrated in Fig. 5A-C.

DETAILED DESCRIPTION OF THE INVENTION

[0026] Bacteria in their natural environment are exposed to many conditions that are not normally experienced in standard industrial or laboratory growth, and thus carry a large number of condition-dependent, stress-induced genes or otherwise nonessential genes which may not be needed in industrial or laboratory use of the organisms. This invention began with the realization that much of the genetic information contained within the genome of a bacteria strain could be deleted without detrimental effect to use of bacteria cultures in processes of industrial or laboratory importance. It was recognized that a bacterium with a reduced genome might be advantageous over native strains in many industrial and laboratory applications. For example, a bacterium with a reduced genome is at least somewhat less metabolically demanding and thus can produce a desired product more efficiently. In addition, a reduced genome can lead to fewer native products and lower level of certain native proteins, allowing easier purification of a desired

protein from the remaining bacterial proteins. Furthermore, some bacterial genetic sequences are associated with instabilities that can interfere with standard industrial or laboratory practices, and might entail costly and burdensome quality control procedures.

[0027] The present invention also involves several methods for deleting genomic DNA from a genome without leaving any inserted DNA behind (scarless deletion). If one is making several sequential deletions from the single DNA molecule which makes up a bacterial genome, it is important not to leave any inserted DNA sequences behind. Such inserted sequences, if they were left behind, would be candidate sites for undesired recombination events that would delete uncharacterized and perhaps important portions of the remaining genome from the bacteria or cause other unanticipated genome rearrangements with untoward effects. Since one of the objectives of the genome reduction effort is to increase the genetic stability of the bacteria, leaving any inserted DNA behind would be contrary to the objective, and should be avoided. Thus the methods used to delete DNA from the genome become important and sophisticated.

[0028] In one aspect, the present invention relates to a bacterium having a genome that is genetically engineered to be smaller than the genome of its native parent strain. For exemplary purposes, the work described here has focused on the common laboratory and industrial bacterium *Escherichia coli*. The genome reduction work described here began with the laboratory *E. coli* strain K-12, which had prior to the work described here, a genome of 4,639,221 nucleotides or base pairs. The bacterium of the present invention can have a genome that is at least two percent (2%), preferably over five percent (5%), more preferably over seven percent (7%) to eight percent (8%) to fourteen percent (14%) to eighteen percent (18%) to twenty percent (20%), to forty percent (40%) to sixty percent (60%) smaller than the genome of its native parental strain. The term "native parental strain" means a bacteria strain (or other organism) found in natural or native environment as commonly understood by the scientific community and on whose genome a series of deletions can be made to generate a bacterial strain with a smaller genome. The percentage by which a genome has become smaller after a series of deletions is calculated by dividing "the total number of base pairs deleted after all of the deletions" by "the total number of base pairs in the genome before all of the deletions" and then multiplying by 100.

[0029] Another aspect of the present invention comprises a reduced genome bacteria in which about 5% to about 10% of its protein coding genes are detailed. Preferably about 10% to 20% of the protein coding genes are deleted. In another embodiment of the invention, about 30% to about 40%)to about 60% of the protein encoding genes are deleted.

[0030] Generally speaking, the types of genes, and other DNA sequences, that can be deleted are those the deletion of which does not adversely affect the rate of survival and proliferation of the bacteria under specific growth conditions. Whether a level of adverse effect is acceptable depends on a specific application. For example, a 30% reduction in proliferation rate may be acceptable for one application but not another. In addition, adverse effect of deleting a DNA sequence from the genome may be reduced by measures such as changing culture conditions. Such measures may turn an unacceptable adverse effect to an acceptable one. Preferably, the proliferation rate is approximately the same as the parental strain. However, proliferation rates ranging from about 5%, 10%, 15%, 20%, 30%, 40% to about 50% lower than that of the parental strain are within the scope of the invention. More particularly, preferred doubling times of bacteria of the present invention may range from about thirty minutes to about three hours.

[0031] The bacteria of the present invention maybe engineered by the methods of the present invention to optimize their use of available resources (*i.e.*, nutrients) for the production of desired products. Those products may be recombinant proteins, by way on non-limiting example insulin, interleukins, cytokines, growth hormones, growth factors, erythropoietin, colony stimulating factors, interferon, antibodies, antibody fragments, or any other useful recombinant protein. The recombinant product may be a therapeutic product, a vaccine component, a diagnostic product, or a research reagent. The bacteria may also be used as a background to express industrially useful products such as commercially useful metabolic intermediates and end products such as vanillin, shikimic acid, amino acids, vitamins, organic acids, and the like, and chemical compounds not naturally produced in the bacteria but produced as a result of metabolic pathway engineering or other genetic manipulation - (see, *e.g.*, U.S. Patent No. 6,472,16 and 6,372,476, both of which are incorporated herein by reference).

[0032] Below, *E. coli* is used as an example to illustrate the genes and other DNA sequences that are candidates for deletion in order to generate a bacterium that can produce a desired product more efficiently. The general principles illustrated and the types of genes and other DNA sequences identified as candidates for deletion are applicable to other bacteria species or strains. It is understood that genes and other DNA sequences identified below as deletion candidates are only examples. Many other *E. coli* genes and other DNA sequences not identified may also be deleted without affecting cell survival and proliferation to an unacceptable level.

[0033] It is assumed in the analysis and methodology described below that at least part of the DNA sequence of the target bacterial strain bacteriophage genome or native plasmid is available. Preferably, the entire sequence is available. Such complete or partial sequences are readily available in the GenBank database. The full genomic sequence of several strains of *E. coli* is, of course, now published (for example, Blattner et al, *Science*, 277:1453-74, 1997 K-12 Strain MG1655; See also GenBank Accession No. U00096; Perna et al, *Nature*, 409, 529-533, 2001; Hayashi et al, *DNA Res.*, 8, 11-22, 2001, and Welch *et al., Proc. Natl. Acad. Sci., USA* (2002) 99 (26) 17020-17024 and GenBank Accession No. AE014075, all of which are incorporated herein by reference in their entirety), as is the sequence of several other commonly used laboratory bacteria. To start the deletion process, the genome of the bacteria is analyzed to look for those sequences that represent good candidates for deletion. Of course, these techniques can also be applied to partially sequenced genomes in the genomic areas for which sequence date is available or could be determined.

[0034] In *E. coli*, and other bacteria as well, as well as in higher organisms, a type of DNA sequence that can be deleted includes those that in general will adversely affect the stability of the organism or of the gene products of that organism. Such elements that give rise to instability include transposable elements, insertion sequences, and other "selfish DNA" elements which may play a role in genome instability. For example, insertion sequence (IS) elements and their associated transposes are often found in bacterial genomes, and thus are targets for deletion. IS sequences are common in *E. coli*, and all of them may be deleted. For purposes of clarity in this document, we use the term IS element and transposable element generically to refer to DNA elements, whether intact or defective, that can move from one point to another in the genome. An example of the detrimental effects of IS elements in science and technology is the fact that they can hop from the genome of the host *E. coli* into a BAC plasmid during propagation for sequencing. Many instance are found in the human genome and other sequences in the GenBank database. This artifact could be prevented by deletion from the host cells of all IS elements. For a specific application, other specific genes associated with genomic instability may also be deleted.

[0035] Shown in Fig. 1 is illustration of the *E. coli* genome, which natively, in the K-12 strain, comprises 4,639,221 base pairs. Fig. 1, shows, on the inner ring, the scale of the base pair positions of the *E. coli* K-12 genome (strain MG1655), scaled without deletions (see also Blattner *et al., supra*). The next ring progressively outward shows regions of the K-12 genome that are missing or highly altered in a related strain O157:H7, and which are thus potentially detectable

from the K-12 genome. The next ring outward shows the positions of the IS elements, both complete and partial, in the native genome. The next ring moving outward shows the positions of the RHS elements A to E and flagellar and restriction regions specially targeted for deletion here. The outermost ring shows the location of the deletions actually made to the genome, as also listed in Tables 1 and 2 below. These deletions make up about 14 percent of the base pairs in the original K-12 MG1655 genome. Using methods of the present invention 18% to 20% to about 40% of the genome will be deleted using the design paradigms described herein.

[0036] Another family of *E. coli* genes that can be deleted are the restriction modification system genes and other endogenous nucleases whose products destroy foreign DNA. These genes are not important for bacterial survival and growth in culture environments. These genes can also interfere with genetic engineering by destroying plasmids introduced into a bacterium. Positions of restriction modification system genes on an *E. coli* genome map are shown in Fig. 1 and Table 1. In one embodiment of the invention, other DNA methylase genes may be added back to the deleted *E. coli* strain so as to optimize the strain for certain uses, for example, eukaryotic methylase genes.

[0037] Another family of *E. coli* genes that can be deleted is the flagella gene family. Flagella are responsible for motility in bacteria. In natural environments, bacteria swim to search for nutrients. In cultured environments, bacteria motility is not important for cell survival and growth and the swimming action is metabolically very expensive, consuming over 1% of the cellular energy to no benefit. Thus, the flagella genes may be deleted in generating a bacterium with a smaller genome. Positions of flagella genes on an *E. coli* genome map are shown in Fig. 1 and Table 1.

[0038] One type of *E. coli* DNA element, already mentioned, that can be deleted is the IS elements (or transposable elements). IS elements are not important for bacteria survival and growth in a cultured environment and are known to interfere with genome stability. Thus, the IS elements can be deleted in generating a bacterium with a smaller genome. Positions of the IS elements on an *E. coli* genome map are shown in Fig. 1 and Table 1.

[0039] Another type of *E. coli* DNA element that can be deleted is the Rhs elements. All Rhs elements share a 3.7 Kb Rhs core, which is a large homologous repeated region (there are 5 copies in *E. coli* K-12) that provides a means for genome rearrangement via homologous recombination. The Rhs elements are accessory elements which largely evolved in some other

background and spread to *E. coli* by horizontal exchange after divergence of *E. coli* as a species. Positions of the Rhs elements on an *E. coli* genome map are shown in Fig. 1 and Table 1.

[0040] One type of region in the *E. coli* genome that can be deleted is the non-transcribed regions because they are less likely to be important for cell survival and proliferation. Another type of regions in the *E. coli* genome that can be deleted is the hsd regions. The hsd regions encode for the major restriction modification gene family which has been discussed above. Positions of the non-transcribed regions and the hsd regions on an *E. coli* genome map are shown in Fig. 1 and Table 1.

[0041] Prophages, pseudogenes, toxin genes, pathogenicity genes, periplasmic protein genes, membrane protein genes are also among the genes that may be deleted, based on the gene selection paradigm discussed herein. After the sequence of *E. coli* K-12 (see Blattner, *et al., supra*), was compared to the sequence of its close relative 0157:H7 (See Perna *et al., supra*) and it was discussed that 22% (K-12) and 46% (O157:H7) of the protein encoding genes were located on strain specific islands of from one to about 85 kb inserted randomly into a relatively constant backbone.

[0042] Among other genes that may be deleted are genes that encode bacteriophage receptors including, for example, ton A (FhuA) and/or its complete operon fhu ABC which encodes the receptor for the lytic phage T1.

[0043] One general method to identify additional genes and DNA sequences as deletion candidates is to compare the genome of one bacterial strain to one or more others strains. Any DNA sequences that are not present in two or three of the strains are less likely to be functionally essential and thus can be used for identifying candidates for deletion. In the examples described below, the complete genomic sequences of two *E. coli* strains, O157:H7 EDL933 and K-12 MG1655, were compared. DNA sequences that were not found in both strains were used to identify targets for deletion. Twelve such identified targets from *E. coli* strain MG1655 were deleted, resulting in a bacteria strain with a genome that is about 8% smaller. The bacteria with the reduced genome grow at substantially the same rate as the native parent MG1655 strain.

[0044] The DNA sequence of a uropathogenic *E. coli* strain CFT073 H7 (see Welch *et al., supra*), was recently determined and its sequence was compared to the K-12 (MG1655) and O157:H7. Results show that only about 40% of all coding genes found in any one of the genomes is present in all of the genomes and CFT073, K-12 and 0157:H7 are composed of 67%, 43% and 68% strain specific island genes. Based on this information, as much as about 60% of the protein

coding sequences may be deleted from *E. coli*. Preferably at least 5% or about 90% or about 15% or about 21% of the protein coding genes are deleted. More preferably, about 30% of the protein coding genes are deleted. It should be noted that there may be genes essential for growth in one strain that are not required for growth in other strains. In such cases, the gene essential for growth of that strain is not deleted from the strain or if deleted is replaced with another gene with a complementary function so as to permit growth of the strain.

[0045] In a particular embodiment of the invention, sequence information is used to select additional genes from (using the methods of the present invention) an *E. coli* genome so as to produce a genome of about 3.7 megabases (about 20% smaller than K-12) containing 73 deletions to remove about 100 "islands" and surrounding DNAs that will still allow for adequate growth of the strain when cultured on minimal media. The design also calls for complete elimination of any remaining transposable elements (IS sequences) from the genome.

Perisplasmic Cleansing and Protein Expression

[0046] For reasons discussed herein, there remains a need in the art for production of recombinant proteins which will be secreted into the periplasmic space of bacteria and the methods of the present invention provide for the engineering of bacteria to optimize periplasmic expression.

[0047] Gram-negative bacteria, such as *E. coli,* have two cellular membranes, the inner cell membrane and the outer cell membrane. Two membranes are separated by a periplasmic space (PS). Bacterial proteins with appropriate signal sequences are secreted through the inner cell membrane into the PS by at least two different systems, Sec-system and Tat-system. (Danese *et al., Annu. Rev. Genet.* (1998) 32:59-94; Fekes *et al., Microbiol. Mol. Biol. Rev.*, (1999) 63: 161-193; and Pugsley, *Microbiol. Rev.* (1993) 57:50-108 [sic]. Hynds *et al.*, (1998) *J. Biol. Chem.* 273:34868-34874; Santini *et al.* (1998) *EMBO J.* 17:101-112; Sargent *et al., EMBO J.* 17:101-112 [TAT] all of which are incorporated herein by reference.

[0048] The Sec-system recognizes an appropriate signal peptide and transports the protein, using cytoplasmic ATP and electronmotive force, into the periplasm in an unfolded state. After cleavage of the signal protein, the new protein folds with the aid of chaperones, peptidyl-prolyl isomerases, and a thioredoxin linked system which catalyses disulfide bond formation. See, *e.g.,* Hynds *et al.*, (1998) *J. Biol. Chem.* 273:34868-34874; Santini *et al.* (1998) *EMBO J.* 17:101-112; Sargent *et al., EMBO J.* 17:101-112 [TAT] all of which are incorporated herein by reference.

[0049] In contrast to Sec-system, the Tat-system transports large proteins in fully folded conformation and is more specific in recognition of appropriate signal sequences. We have selected the periplasm because (1) it is a preferred site for expressing heterologous recombinant proteins, (2) for industrial use in controlled conditions, it has many unnecessary proteins, and (3) it plays a role in many unnecessary adaptation and control systems, some of which appear to be detrimental. By removing native proteins from the periplasm, we anticipate that we will be able to greatly improve the process for protein production. Expression and secretion of proteins in the periplasm has been reviewed in *Hanahan, D., J. Mol. Biol., 1983, 166(4):p. 557-80; Hockney, R.C., Trends Biotechnol., 1994, 12(11):p. 456-632.;* and *Hannig G., et al., Trends Biotechnol., 1998, 16(2): p. 54-60.* all of which are incorporated by reference.

[0050] There are several reasons why the periplasm is a preferred site for protein production; (1) it is possible to produce a recombinant protein with the amino terminus identical to the natural protein, whereas in the cytoplasm, proteins invariably begin with the amino acid methionine; (2) many proteins can fold correctly in the periplasmic space (3) the correct disulfide bonds can form in the oxidising environment of the periplasm; (4) the periplasmic space contains much less and far fewer proteins than the cytoplasm, simplifying purification (5) there are fewer proteases than in the cytoplasm, reducing protein digestion and loss; (6) expressed proteins can be readily released with other periplasmic proteins by specifically disrupting the outer membrane, substantially free of the more abundant cytoplasmic proteins. The periplasmic space has natural enzyme systems, linked to cellular cytoplasmic metabolism through the inner membrane, to undertake these processing tasks, presumably because this is the organelle in which most inner and outer membrane proteins are processed. By contrast, it has proven very difficult to obtain proper folding of recombinant protein chains expressed in the reducing environment of the cytoplasm. Often proteins aggregate into insoluble "inclusion bodies." Whilst initial inclusion body purification might be simpler, the proteins need to be re-dissolved and re-folded, a process that is unpredictable and difficult to control, and for some proteins, so inefficient as to be unworkable at industrial scale.

[0051] Recombinant proteins are generally produced in the periplasm by expressing fusion proteins in which they are attached to a signal peptide that causes secretion into the periplasmic space. There the signal peptide is cleaved off very precisely by specific signal peptidases. Second generation recombinant human growth hormone is manufactured by this method by Genentech (*Nutropin, Full Prescriber Information.*) and Pharmacia. Not all proteins can be successfully

Biology (1994) 16.6.1-16.6.14 (Copyrighted 2000 by John Wiley *et al.* and Sons) all of which are incorporated herein by reference in their entirety.

[0055] In one embodiment of the present invention, nine known and 3 putative periplasmic protein genes were successfully deleted in constructing MDS40, without significantly affecting the ability of the organism to grow on minimal medium. (See Table 4 and data below). These mutations affect a range of functions, including amino acid uptake, inorganic metabolism, cell membrane maintenance, sugar metabolism, and adhesion.

[0056] Approximately 85 genes have been deleted that code for known or putative membrane proteins, identified by their signal-peptide sequences. Of these 33 are involved in flagellar structure or biosynthesis; 9 are involved in fimbrial structure or biosynthesis; and 13 are involved in general secretory pathways. The remainder have a variety of known or putative functions in the cell membranes. Many of these proteins are believed to be processed in the periplasmic space. They have also been deleted in constructing MDS40, without significantly affecting the ability of the organism to grow on minimal medium.

[0057] By searching for signal peptide-like sequences in annotated MG1655 databases, and cross-relating these with the literature we have identified 181 proteins that the majority of which are believed to be resident periplasmic proteins. A number of these proteins have been classified according to function into several groups excluding: adhesion and mobility; nutrient and salt uptake, trace element uptake; environmental sensing; defense and protection; and periplasmic protein secretion and processing. Among the genes or full operons which have been or will be deleted are those coding for sugar and amino acid transport proteins, unlikely to be needed in defined minimal media say for biopharmaceutical production.

[0058] To monitor efficiency of the recombinant protein transportation into PS, either of three commercially available tags: *E. coli* alkaline phosphatase, *Aequoria* green fluorescent protein (GFP) or human growth hormone protein may be used according to the methods described above. The human growth hormone protein is currently most preferable for final demonstration purposes and will be used in ELISA and gene chip-based measurements of the recombinant protein localization to PS.

[0059] One can test the consequence of deleting one or several genes or other DNA sequences from the genome. For example, after one or several genes or other DNA sequences of the genome have been deleted, one can measure the survival and proliferation rate of the resultant bacteria. Although most of the above-identified genes or other DNA sequences may be deleted

without detrimental effect for purpose of producing a desired product, it is possible that the deletion of a specific gene or other DNA sequence may have an unacceptable consequence such as cell death or unacceptable level of reduction in proliferation rate. This possibility exists because of redundancies in gene functions and interactions between biological pathways. Some deletions that are viable in a strain without additional deletions will be deleterious only in combination with other deletions. The possibility exists also because of certain methods used to identify deletion candidates. For example, one method used to identify deletion candidates is to compare two *E. coli* strains and select genes or other DNA sequences that are not present in both strains. While the majority of these genes and other DNA sequences are not likely to be functionally essential, some of them may be important for a unique strain. Another method used to identify deletion candidates is to identify non-transcribed regions and the possibility exists that certain non-transcribed regions may be important for genome stability.

[0060] The consequence of deleting one or several genes or other DNA sequences to be tested depends on the purpose of an application. For example, when high production efficiency is the main concern, which is true for many applications, the effect of deletions on proliferation rate and medium consumption rate can be the consequence tested. In this case, the consequence tested can also be more specific as the production speed quantity and yield per cell of a particular product. When eliminating native protein contamination is the main concern, fewer native proteins and lower native protein levels, or the absence of a specific native protein, can be the consequence tested.

[0061] Testing the consequence of deleting a gene or other DNA sequence is important when little is known about the gene or the DNA sequence. Though laborious, this is another viable method to identify deletion candidates in making a bacterium with a reduced genome. This method is particularly useful when candidates identified by other methods have been deleted and additional candidates are being sought.

[0062] When the consequence of deleting a gene or other DNA sequence has an effect on the viability of the bacteria under a set of conditions, one alternative to not deleting the specific gene or other DNA sequence is to determine if there are measures that can mitigate the detrimental effects. For example, if deleting lipopolysaccharide (LPS) genes results in poor survival due to more porous cellular membranes caused by the absence from the cellular membranes of the transmembrane domain of the LPS proteins, culture conditions can be changed to accommodate

the more porous cellular membranes so that the bacteria lacking the LPS genes can survive just as well as the bacteria carrying the LPS genes.

[0063] Methods for deleting DNA sequences from bacterial genomes that are known to one of ordinary skill in the art can be used to generate a bacterium with a reduced genome. Examples of these methods include but are not limited to those described in Posfai, G. et al., *J. Bacteriol.* 179: 4426-4428 (1997), Muyrers, J.P.P. et al., *Nucl. Acids Res.* 27:1555-1557 (1999), Datsenko, K.A. et al., *Proc. Natl. Acad. Sci.* 97:6640-6649 (2000) and Posfai, G. et al., *Nucl. Acids Res.* 27: 4409-4415 (1999), all of which are hereby incorporated by reference in their entirety. Basically, the deletion methods can be classified to those that are based on linear DNAs and those that are based on suicide plasmids. The methods disclosed in Muyrers, J.P.P. et al., *Nucl. Acids Res.* 27:1555-1557 (1999) and Datsenko, K.A. et al., *Proc. Natl. Acad. Sci.* 97:6640-6649 (2000) are linear DNA-based methods and the methods disclosed in Posfai, G. et al., *J. Bacteriol.* 179: 4426-4428 (1997) and Posfai, G. et al., *Nucl. Acids Res.* 27: 4409-4415 (1999) are suicide plasmid-based methods.

[0064] Some known methods for deleting DNA sequences from bacterial genomes introduce extraneous DNA sequences into the genome during the deletion process and thus create a potential problem of undesired homologous recombination if any of the methods is used more than once in a bacterium. To avoid this problem, scarless deletion methods are preferred. By scarless deletion, we mean a DNA sequence is precisely deleted from the genome without generating any other mutations at the deletion sites and without leaving any inserted DNA in the genome of the organism. However, due to mistakes, such as those made in PCR amplification and DNA repairing processes, one or two nucleotide changes may be introduced occasionally in scarless deletions. Described below are some novel scarless deletion methods, either linear DNA-based or suicide plasmid-based. These novel methods have been applied to *E. coli* strains in the examples described below. It is understood that the specific vectors and conditions used for *E. coli* strains in the examples can be adapted by one of ordinary skill in the art for use in other bacteria. Similar methods and plasmids can be used to similar effect in higher organisms. In some instances it may be more appropriate to modify an existing production strain rather than transfer production to the minimized genome *E. coli* strain.

[0065] The methods of the present invention are not limited to use in reducing the genome of bacteria, for example, the present methods may be used to delete DNA from bacteriophage such as P1, P2, lambda and other bacteriophage. Such methods permit the engineering of

bacteriophage genomes so as to improve their useful properties and/or to decrease or eliminate certain properties which impair the use of such bacteriophage for a variety of purposes. Similarly, the methods of the present invention are useful for modifying plasmids that reside in bacteria so as to eliminate harmful elements (*e.g.*, virulence genes) from the plasmid and to improve other useful properties of the plasmids.

[0066] The well known generalized transducing bacteriophage P1 has been as described above for transducing pieces of DNA into recipient *E. coli.* Certain gene features of P1, however, ultimately limit the capacity to pick up and package genomic DNA for transduction. In particular, the packaging site (pac) site of P1 is a GATC rich region which when methylated by the *dam* methylase of P1 limits the amount of genomic DNA into the phage coat. However in the absence of ***dam*** associated methylation of the packaging site, packaging of DNA becomes "sloppy", that is, it more readily packages portions of genomic DNA than would be the case if the packaging site were methylated. Therefore, it would be advantageous to engineer the P1 genome to remove *dam* gene using the deletion methods of the present invention thereby enhancing the ability to pick up and package genomic material. *dam*

[0067] Another drawback associated with the use of P1 transduction in that the phage carries two insertion sequences. On insertion sequence, IS1 is found between *ssb* and the *prt* loci of the P1 genome. Another, IS5 is in the *res* gene As a result, it is possible that when P1 is used in transduction that one or more of the insertion sequences could end up jumping into a genomic locus of the organism. Therefore, it would be advantageous to engineer the P1 genome to delete the IS sequences using the methods of the present invention thereby preventing genomic contamination where P1 is used as a transduct.

[0068] In the above description, the present invention is described in connection with specific examples. It will be understood that the present invention is not limited to these examples, but rather is to be construed to be of spirit and scope defined by the appended claims.

[0069] Among the embodiments of the present invention is a *Shigella flexneri* having a reduced genome. Recently, the complete genome sequence of *Shigella flexneri* 2a strain 2457T was determined. (The sequenced strain was redeposited at the American Type Culture Collection, as accession number ATCC 700930.) The genome of *S. flexneri* consists of a single-circular chomosome of 4,599,354 base pairs (bp) with a G+C content of 50.9%. Base pair 1 of the chromosome was assigned to correspond with base pair one of *E. coli* K-12 since the bacteria shows extensive homology. The genome was shown to about 4082 predicted genes with an

average size of 873 base pairs . The *S. flexneri* genome exhibits the backbone and island mosaic structure of *E. coli* pathogens albeit with much less horizontally transferred DNA and lacks 357 genes present in *E. coli*. (*See,* Perna *et al.*, (2001) *Nature, 409:529-533*. The organism is distinctive in its large complement of insertion sequences, several genomic rearrangements, 12 cryptic prophages, 372 pseudogenes, and 195 *Shigella* specific genes. The completed annotated sequence of *S. flexneri* was deposited at GenBank accession number AE014073 which is incorporated herein by reference. (See also "Complete Genome Sequence and Comparative Genomics of *Shigella flexneri* Serotype 2A strain 2457T", Wei *et al.*, submitted for publication.) It is striking to note that based on its DNA sequence, *Shigella* is phylogenetically indistinguishable from *E. coli*.

[0070] As is readily apparent from this disclosure, having the *S. flexneri* sequence in hand, its genome may be readily reduced using the methods and gene selection paradigms discussed herein. A reduce genome *Shigella* may be useful for the expression of heterologous (recombinant) proteins or other useful nutrients for reasons discussed herein with respect to reduced genome *E. coli* (live vaccine). Another use for reduced genome *Shigella* or for that matter any pathogenic bacteria susceptible to the deletion methods of the present invention is as a vehicle for the display or presentation of antigens for the purpose of inducing an immune response from a host. Such an engineered *Shigella* could, for example, have genes responsible for virulence deleted from the organism while maintaining other genes such as those encoding antigenic determinants sufficient to induce an immune response in a host and preferably a mucosal immune response in the intestinal wall of a host.

[0071] *Shigella flexneri* is potentially well suited for this strategy in that its virulence determinants have been characterized and have been localized to a 210-kb "large virulence (or Invasion) plasmid" whose nucleotide sequence has been determined and has been deposited as GenBank Accession No. AF348706 which is incorporated herein by reference. (See also Venkatesan *et al. Infection of Immunity* (May 2001) 3271-3285). Among the likely candidates for deletion from the Invasion plasmid is the *cadA* gene which encodes lysine decarboxylase.

[0072] The deleted *Shigella* invasion plasmid may be introduced into a reduced genome *E. coli* thereby allowing efficient expression of certain *Shigella* invasion plasmid genes capable of giving rise to an immune response in a host inoculated with the *E. coli*. The invasion plasmid may also be engineered to delete harmful genes from the plasmid such as the genes responsible for vacuole disruption. Preferred candidate genes for removal from the invasion plasmid include one

or more genes selected from the group consisting of ipaA, ipaB, ipaC, ipaD and virB. The present invention also allows the addition of other genes to the reduced genome-*E. coli* into which the invasion plasmid has been introduced so as to optimize expression of genes from the introduced, modified invasive plasmid.

[0073] The present invention is also directed to live vaccines comprising a reduced genome, for example, *E. coli*, or a reduced genome, for example, *E. coli* into which has been introduced genes encoding antigens capable of inducing an immune response in a host who has been inoculated with the vaccine. Reduced genome vaccines may be DNA based vaccines in containing a DNA known to be capable of inducing a desired physiological response in a hose (i.e., immune response).

[0074] One of the major advantages of a reduced genome organism according to the present invention is to provide a clean, minimal genetic background into which DNAs may be introduced to not only allow expression of a desired molecule, but it also affords the opportunity to introduce additional DNAs into the clean background to provide a source of molecules capable of optimizing expression of the desired product.

Deletion Methods

Construction of a linear targeting DNA

[0075] An example of the construction of a linear target DNA is as follows: To generate primer a+b (Fig. 1), 20 pmol of primer **a** was mixed with 20 pmol of primer **b** and PCR was performed in a total volume of 50 μl. Cycle parameters were: 15 x (94°C 40sec/57°C or lower[depending on the extent of overlap between primers **a** and **b**] 40 sec/72°C 15 sec). Next 1μl of this PCT product was mixed with 20 pmol of primers **a** and **c** (Fig. 1) each, 50 ng of pSG76-CS template and a second round of PCT was performed in a volume of 2 x 50 μl. cycle parameters were: 28x(94°C 40sec/57°C 40sec/72°C 80sec). The resulting, PCR-generated linear DNA-fragment was purified by Promega Wizard PCT purification kit, and suspended in 20 μl water. Elimination of the template plasmid (*e.g.*, by DpnI digestion) is not needed. **pSG76-CS** serves as a template plasmid to generate linear targeting fragments by PCT. It contains the chloramphenicol resistance (Cm^R) gene and two I-*Sce*I sites, and was obtained by the PCT-mediated insertion of a second I-*Sce*I sites, and was obtained by the PCT-mediated insertion of a second I-*Sce*I recognition site into pSG76-C, downstream of the *Not*I site. The two I-*Sce*I sites are in opposite orientation.

Novel linear DNA-based scarless deletion method I

[0076] The novel DNA-based scarless deletion method of the present invention can be best understood when the following description is read in view of Fig. 2. Generally speaking, the method involves replacing a segment of the genome, marked for deletion, with an artificial DNA sequence. The artificial sequence contains one or more recognition sites for a sequence-specific nuclease such as I-SceI, which cuts at a sequence that does not occur natively anywhere in the *E. coli* K-12 genome. Precise insertion of the linear DNA molecule into the genome is achieved by homologous recombination aided by a system that can increase the frequency of homologous recombination. When the sequence-specific nuclease is introduced into the bacteria, it cleaves the genomic DNA at the unique recognition site or sites, and only those bacteria in which a homologous recombination event has occurred will survive.

[0077] Referring specifically to Fig. 2, the plasmid pSG76-CS is used as a template to synthesize the artificial DNA insert. The artificial insertion sequence extends between the sequences designated A, B and C in Fig. 2. The C^R indicates a gene for antibiotic resistance. The insert DNA is PCR amplified from the plasmid and electroporated into the *E. coli* host. The insert was constructed so that the sequences A and B match sequences in the genome of the host which straddle the proposed deletion. Sequence C of the insert matches a sequence in the host genome just inside sequence B of the host genome. Then the bacteria are selected for antibiotic resistance, a selection which will be survived only by those bacteria in which a homologous recombination event occurred in which the artificial DNA inserted into the bacterial genome. This recombination event occurs between the pairs of sequences A and C. The inserted DNA sequence also includes a sequence B, now positioned at one end of the insert, which is designed to be homologous to a sequence in the genome just outside the other end of the insert, as indicated in Fig. 2. Then, after growth of the bacteria, the bacteria is transformed with a plasmid, pSTKST, which expresses the I-SceI sequence-specific nuclease. The I-SceI enzyme cuts the genome of the bacteria, and only those individuals in which a recombination event occurs will survive. 10-100% of the survivors are B to B recombination survivors, which can be identified by a screening step. The B to B recombination event deletes the entire inserted DNA from the genome, leaving nothing behind but the native sequence surrounding the deletion.

[0078] To repeat, the first step of the method involves providing a linear DNA molecule in a bacterium. The linear DNA molecule contains an artificial linear DNA sequence that has the following features: one end of the linear DNA sequence is a sequence identical to a genome

sequence on the left flank of the genome region to be deleted, followed by a sequence identical to a genome sequence on the right flank of the genome region to be deleted; the other end of the linear DNA molecule is a sequence identical to a genome sequence within the genome region to be deleted; between the two ends of the linear DNA, there is a recognition site that is not present in the genome of the bacterial strain and an antibiotic selection gene. The artificial DNA sequence can be made using polymerase chain reaction (PCR) or directed DNA synthesis. A PCR template for this purpose contains the unique recognition site and the genomic DNA sequences on both ends of the artificial linear DNA sequence are part of the primers used in the PCR reaction. The PCR template can be provided by a plasmid. An example of a plasmid that can be used as a template is pSG76-C (GenBank Accession No. Y09893), which is described in Posfai, G. et al., *J. Bacteriol.* 179: 4426-4428 (1997). pSG76-CS (GenBank Accession No. AF402780), which is derived from pSG76-C, may also be used. pSG76-CS contains the chloramphenicol resistance (Cm^R) gene and two I-SceI sites, and was obtained by the PCR-mediated insertion of a second I-SceI recognition site into pSG76-C, downstream of the NotI site. The two I-SceI sites are in opposite direction.

[0079] An artificial or constructed DNA sequence can be provided to a bacterium by directly introducing the linear DNA molecule into the bacterium using any method known to one of ordinary skill in the art such as electroporation. In this case, a selection marker such as an antibiotic resistance gene is engineered into the artificial DNA sequence for purpose of selecting colonies containing the inserted DNA sequence later. Alternatively, a linear DNA molecule can be provided in a bacterium by transforming the bacterium with a vector carrying the artificial linear DNA sequence and generating a linear DNA molecule inside the bacterium through restriction enzyme cleavage. The restriction enzyme used should only cut on the vector but not the bacterial genome. In this case, the artificial linear DNA sequence does not have to carry a selection marker because of the higher transformation efficiency of a vector so that a bacterium with the inserted linear DNA can be screened by PCR later directly.

[0080] The second step of the scarless deletion method involves replacement of a genomic region by insertion of the artificial DNA molecule. The bacterial cells are engineered to contain a system that increases the frequency of homologous recombination. An example of such a system is the Red recombinase system. The system can be introduced into bacterial cells by a vector. The system helps the linear DNA molecule to replace a genomic region which contains the deletion target. As described in the examples below, a vector carrying a homologous recombination

system that can be used in *E. coli* is pBADαβγ, which is described in Muyrers, J.P.P. et al., *Nucl. Acids Res.* 27:1555-1557 (1999). Another plasmid pKD46 described in Datsenko, K.A. et al., *Proc. Natl. Acad. Sci.* 97:6640-6649 (2000) may also be used. Other plasmids that can be used include pGPXX and pJGXX. pGPXX is derived from pBADαβγ by replacing the origin of replication in pBADαβγ with pSC101 origin of replication. pJGXX is a pSC101 plasmid that encodes the Red functions from phage 933W under tet promoter control

[0081] The third step of the scarless deletion method involves removal of the inserted DNA sequence. An expression vector for a sequence-specific nuclease such as I-SceI that recognizes the unique recognition site on the inserted DNA sequence is introduced into the bacteria. The sequence-specific nuclease is then expressed and the bacterial genome is cleaved. After the cleavage, only those cells in which homologous recombination occurs resulting in a deletion of the inserted linear DNA molecule can survive. Thus, bacteria with a target DNA sequence deleted from the genome are obtained. Examples of sequence-specific nuclease expression vectors that can be used in *E. coli* include pKSUC1, pKSUC5, pSTKST, pSTAST, pKTSHa, pKTSHc, pBADSce1 and pBADSce2. The sequence-specific nuclease carried by these vectors is I-SceI. pKSUC1, pKSUC5, pSTKST and pSTAST are described below in the examples.

[0082] The method described above can be used repeatedly in a bacterium to generate a series of deletions. When the expression vector for the homologous recombination system and the expression vector for the unique sequence-specific nuclease are not compatible with each other, such as the case for pBADαβγ and pKSUC1, transformation of the two vectors have to be performed for each deletion cycle. Transformation of the two vectors can be avoided in additional deletion cycles when two compatible plasmids, such as pBADαβγ and pSTKST, or pKD46 and pKSUC5, are used. An example of using two of these vectors that are compatible with each other is described in the examples below.

[0083] The above scarless deletion method can be modified to make a series of deletions on a bacterial genome more efficient (an example of which is Procedure 4 in Examples below). The first step of the modified method involves making insertions of a linear DNA molecule individually in bacterial cells, preferably wild-type bacteria cells, in a parallel fashion, resulting in a set of strains, each carrying a single insertion. This step can be carried out as described above. The second step of the modified method involves sequentially transferring individual insertions into the target cell whose genome is to be reduced. P1 transduction is an example of the methods

that can be used for transferring insertions. The third step of the modified method involves recombinational removal of the inserted sequence, which can be carried out as described above.

Novel linear DNA-based scarless deletion method II

[0084] In this novel linear DNA-based method, two DNA sequences, one of which is identical to a sequence that flanks one end of a bacterial genome region to be deleted and the other of which is identical to a sequence that flanks the other end of the bacterial genome region and oriented similarly, are engineered into a plasmid vector. The vector is herein termed the target vector. The two DNA sequences are located next to each other on the target vector. At least one recognition site for an enzyme that will only cut the target vector but not the bacterial genome is also engineered into the target vector at a location outside the two DNA sequences. The recognition site can be one for a sequence-specific nuclease such as I-SceI. The recognition site can also be one for a methylation-sensitive restriction enzyme that only cuts an unmethylated sequence. Since the recognition site, if there is any, on the bacterial genome is methylated, the restriction enzyme can only cut the target vector. The target vector is transformed into a bacterium and a linear DNA molecule is generated inside the bacterium by expressing in the bacterium the enzyme that recognizes and cuts the recognition site on the target vector. Next, a system that can increase homologous recombination is activated inside the bacterium to induce homologous recombination between the homologous sequences of the linear DNA and the bacterial genome that flank the region to be deleted. A bacterium with a targeted genome region deleted can be obtained as a result of the above homologous recombination.

[0085] This novel linear DNA-based method can also be used to replace a region of a bacterial genome with a desired DNA sequence. In this case, a desired DNA sequence that can undergo homologous recombination with the bacterial genome to replace a region on the genome is engineered into the target vector. All other aspects are the same as described above for deleting a region of the bacterial genome.

[0086] Regardless whether the method is used to delete or replace a target region in the bacterial genome, a marker gene for selecting incorporation of DNA carried on the target vector into the bacterial genome is not necessary due to the high incorporation efficiency. Simply screening 30-100 colonies by PCR usually allows the identification of a clone with desired modification in the bacterial genome.

[0087] As a specific example, Figs. 3 and 4 illustrates using this method for introducing an Amber stop codon in the middle of a gene. As a first step, a DNA fragment with the desired